

# Sitanshu Kushwaha

✉ sak9813@nyu.edu 📞 929-643-4480 🔗 sitanshu.tech 📍 New York, NY

🌐 linkedin.com/in/sitanshukushwaha/ 🏠 github.com/Sitanshuk

## 🎓 EDUCATION

**New York University**, MS in Computer Science Sep 2023 – May 2025 | New York  
Relevant Courses: Big Data, Data Science, Data Management and strategy, Cloud Computing and Big Data

**University of Mumbai**, BE in Computer Engineering Aug 2016 – Nov 2020 | Mumbai

## 🏢 PROFESSIONAL EXPERIENCE

**Data Engineering Intern**, NBCUniversal Jun 2024 – present | New York

- Implemented a year-round aggregation strategy in **BigQuery** for Peacock's annual user insights, processing **petabyte-scale** data incrementally to distribute computational load, reducing end-of-year query times from **hours to minutes** for **35M+** subscribers, and optimizing resource utilization and costs.
- Reengineered legacy processes by centralizing data in **Databricks' Unity Catalog** and introducing a **self-service tool**, enhancing **data governance** and improving **stakeholder transparency**. This streamlined **forecast extrapolation** across demographics, reducing communication overhead.
- Optimized** data workflows, accelerating **report delivery time** by **2 days** and eliminating **20 hours** of **manual intervention** per quarter. This improved **efficiency** and **accuracy** of forecasting processes.

**Data Engineer**, Enterprise Data Management - NYU IT Oct 2023 – May 2024 | New York

- Implemented 10+ **ELT** pipeline with **Snowflake & AWS**, integrating diverse data sources (Data Warehouses & Data Marts) for enhanced analysis & decision-making.

**Data Engineer**, LTIMindtree

Jan 2021 – Jun 2023 | Mumbai

**Technical Lead**, Visioncare MFF Data Engineering team - Johnson and Johnson

- Optimized **Databricks Spark** code, achieving a **30% reduction in execution time** for 50% of transformation jobs, enhancing data timeliness and **scalability** for **multi-TB datasets**.
- Implemented **event-based triggers** in **Azure Data Factory** for ETL pipelines, enhancing efficiency in handling **Big Data** from diverse sources and reducing **cloud costs by 25%**.
- Designed a **Monitoring Dashboard** in **Tableau** for **real-time data flow architecture**, enabling early identification of bottlenecks and reducing system outages by **40%**.
- Streamlined deployment by implementing **CI/CD** using **Git** and **ARM templates** (Infrastructure as Code) in **Azure**, reducing deployment time by **70%** and ensuring rapid, automated updates to **cloud infrastructure**.

## 🧠 SKILLS

**Big Data** — PySpark, Kafka, Databricks, BigQuery, Snowflake, AWS, GCP, Data Lake, ETL, NoSQL, Airflow,

**Machine Learning** — Scikit Learn, Tensorflow, NLP, Neural Networks, Deep Learning, **Data Analytics** — SQL, Pandas, Numpy, Matplotlib, Web Scraping, Tableau, **Languages** — Python, JAVA, R., **Tools** — Git, Docker

## 📁 PROJECTS

**DineSync - Real-Time Culinary Exploration in NYC**, (Big Data, Spark, Kafka, MongoDB, Django) 🔗

- Engineered DineSync, a **real-time restaurant recommendation** system leveraging **Kafka** for **processing live** user check-ins, ensuring accurate seat availability data with 95% accuracy.
- Engineered a solution that automatically recommended alternative restaurants when primary choices were fully booked, resulting in a **20% decrease in drop-off rates** during peak reservation times.

**Talk2Doc - (GCP, RAG, LLM, APIs)** 🔗

- Architected a centralized ecosystem for students using Google Cloud Platform (**GCP**), integrating Retrieval-Augmented Generation (**RAG**) for personalized note searching and automated job application tracking.
- Architected **modular**, event-driven highly scalable system utilizing **serverless** functions, queues, LLM capable of handling millions of users.

**Subjective Answer Evaluation using Machine Learning**, (NLP, Django, TensorFlow) 🔗

- Incorporated state-of-the-art **NLP** techniques, including **BERT**, **USE**, and **Word2Vec** language models, to assess students' subjective answers by measuring **semantic similarity** against the teacher's answer.
- Published a **research paper** in International Journal for Scientific Research and Development 🔗 .

## 🏆 AWARDS

1st place - **Innovation Business Case Project**, NBCUniversal

Aug 2024

2nd place - **Build with AI**, Google Developer Group NYC

Apr 2024